AD-A116 427 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
IMPROVING THE VALIDITY OF A CRITERION-REFERENCED, DIAGNOSTIC TE--ETC(U)
MAR 81 R M BEARDEN

UNCLASSIFIED NPRDC-TN-81-7 NL

1 OF 1
AD A
116427

END
DATE
FILMED
07-82
DTIC

NPRDC TN 81-7

MARCH 1981

IMPROVING THE VALIDITY OF A CRITERION-REFERENCED, DIAGNOSTIC TEST USING A DISCRIMINANT FUNCTION PROCEDURE

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN DIEGO, CALIFORNIA

NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

NPRDC Technical Note 81-7 March 1981

IMPROVING THE VALIDITY OF A CRITERION-REFERENCED, DIAGNOSTIC TEST USING A DISCRIMINANT FUNCTION PROCEDURE

Ronald M. Bearden

Reviewed by
Gerald Laabs

Released by
Richard C. Sorenson
Director of Programs

Navy Personnel Research and Development Center
San Diego, California 92152

FOREWORD

This research was conducted in support of Project ZF63-522-002-03.40, Techniques for the Measurement of Job Performance. The statistical procedure described in this report was used successfully to improve the overall predictive validity of the scores on the Basic Mechanical Procedures Test, a criterion-referenced, diagnostic test for boiler technicians.

Special appreciation is expressed to V. A. Reisenleiter, Army Research Institute, who brought this statistical procedure to our attention.

RICHARD C. SORENSON
Director of Programs

FL-182 on file

A

SUMMARY

Problem

The development of a criterion-referenced, diagnostic test for boiler technicians (BTs) was a necessary part of the Personnel Readiness Training Program, a diagnostic testing/shipboard training program. In the absence of a well-defined methodology for the optimal weighting of individual item scores on criterion-referenced tests, a systematic and practical procedure leading to the improvement of the predictive validity of scores on this diagnostic test was needed.

Purpose

The purpose of this effort was to determine whether an application of discriminant function analysis could significantly improve the overall predictive validity of the scores on a diagnostic criterion-referenced test. In this statistical procedure, the examinee's score is based on the sum of the individual item discriminant weights as opposed to one for a correct response and zero for an incorrect one.

Method

The test employed in this research was the criterion-referenced, Basic Mechanical Procedures (BMP) test, which was developed by NAVPERSRANDCEN to diagnose individual deficiencies within the BT rating. The sample consisted of 200 BTs assigned to the shore-based Propulsion Engineering School, Service School Command, Great Lakes. Half--the pre-instruction group--were entering the modularized, self-paced curriculum, and half--the post-instruction group--had completed it.

Using the discriminant function procedure, an optimal weighting strategy of the individual item scores was derived for 12 of the 14 modules of the BMP test. This discriminant scoring procedure was then compared to the traditional number-correct procedure in terms of the predictive validity levels within each module. The validity level was estimated by comparing actual group membership (pre- vs. post-instruction) with group membership assigned on the basis of the discriminant and number-correct scores.

Results

1. Most of the modules under both scoring methods demonstrated excellent classification ability with each of the percent agreements significant at conventional levels.

2. A significant improvement in the overall validity of the discriminant scores as compared to number correct was noted.

3. Differences between pairs of validity coefficients in four of the five modules having the poorest classification percentages were significant in favor of the discriminant scores.

4. Scores resulting from the discriminant weights improved the ability of this criterion-referenced test to accurately classify students as members of either the pre- or post-instruction group.

Conclusions

The discriminant function:

1. Would appear to be sound by analogy to statistical theory and would utilize well known item statistics.

2. Allows the items to be scaled along a discrimination continuum with meaningful end points.

3. Provides an index of the usefulness of the items for discriminating between pre- and post-instruction member's test scores. Such a determination of item weights would be easy to program.

4. Should prove especially helpful in improving the overall validity of tests including items that are not as valid as those included in the BMP test.

Recommendation

The discriminant function should be considered as an alternative procedure to the more conventional "number-correct" procedure for determining item scores.

CONTENTS

INTRODUCTION

Problem and Background

Criterion-referenced measurement has been one of the most provocative ideas to influence educational measurement theory and practice in recent years. Although the large number of articles that have been published in this area ot measurement describe a variety of criterion-referenced tests, it is still difficult to find the precise definition of a criterion-referenced test. Typically, criterion-referenced measures are used to provide information on the status of an individual's knowledge and skill with respect to some criterion or standard of performance (Popham & Husek, 1969). Depending on how the test results are used, however, the same test can be either norm- or criterion-referenced (Hambleton & Novick, 1973). For example, using the results of a typing test to choose the fastest typist or assign the highest typing grade represents a norm-referenced procedure. Using the results of the same typing test to decide that more practice or training is needed when a cutoff score is not met represents a criterion-referenced procedure. The latter application contains the essence of the definition of a criterion-referenced test that will be used in this note. That is, criterion-referenced measurement emphasizes the description of the absolute rather than relative level of performance with respect to a well-defined behavior domain.

In the context of the above definition of criterion-referenced measurement, test validity has been approached in the framework of both a modified classical theory (Livingston, 1972) and Bayesian statistics (Hambleton & Novick, 1973). In each instance, classical procedures for estimating variability of criterion-referenced test scores yielded spuriously low validity estimates based on correlational procedures. In terms of the purpose of these tests, which is to determine the degree to which any student has mastered a set of objectives, either face or content validity measures are considered appropriate solutions to the validation problem. The content validity approach, which is

better suited for such tests, can be determined, according to Popham and Husek (1969), by "a carefully made judgment, based on the test's apparent relevance to the behaviors legitimately inferable from those delimited by the criterion." If a technique such as advocated by Bormuth (1970) for defining content domains and item generation rules is followed, content validity is necessarily guaranteed.

Inherent in the qualitative approaches to the measurement of test validity is the inability to quantify the degree to which criterion-referenced tests lead to accurate diagnostic decisions. Fortunately, Panell and Laabs (1979) developed an alternative procedure that does provide a useful, quantitative estimate of the predictive validity level of criterion-referenced test items. In this procedure, students in a cross-validation sample are classified as either mastery or nonmastery achievers, according to how their performance compared to a predetermined cutoff score. This classification is then compared with actual group membership, and the percent agreement between predicted and actual group membership is obtained. Once an estimate is derived using this procedure, a search for alternative scoring methods that might further enhance the reliability and validity of criterion-referenced items can be made.

For more than three decades, researchers have investigated the effects of scoring methods on reliability and validity of norm-referenced test items. A small sample of the many variations appearing in the literature include elimination scoring (Coombs, Milholland, & Womer, 1956), Guttman weighting (Guttman, 1941), judged confidence weighting (Patnaik & Traub, 1973), confidence weighting (Shuford, Albert, & Massengill, 1966), and the conventional correction for guessing (Lord & Novick, 1968). In each of these methods, an attempt is made to increase the reliability and validity of inferences made from scores by capitalizing on the different levels of knowledge reflected in both the correct and incorrect responses. For example, in the judged confidence method, the responses to a multiple-choice item are differentially weighted according to their

prejudged degree of correctness. In the elimination method, the examinee is instructed to indicate which of the options he can identify as incorrect, and his score is determined by the number of distractors correctly identified. In confidence weighting, the response alternative the examinee would have selected is inferred from the manner in which he has assigned personal probabilities to the item-choices. The question of whether the use of scoring procedures other than number-correct does lead to more valid tests, despite the intuitive appeal of differential response weighting, has not been clearly answered. In point of fact, Stanley and Wang (1970), in their comprehensive review of differential response weighting, report that studies of this kind have been far from conclusive in demonstrating consistent gains in desired psychometric properties. In addition, since these methods do not fit within the criterion-referenced measurement framework, they should not be used for improvement of such tests.

An alternative procedure to the differential weighting of response options used with norm-referenced tests, one that is acceptable for criterion-referenced measurement usage, involves optimal weighting of the individual item scores themselves in an effort to maximize the overall predictive validity of the test. A possible methodological candidate capable of providing this type of optimal weighting strategy is the plug-in discriminant function analysis for discrete binary items (Elvers, 1977).[1] In this statistical procedure, which was brought to our attention by V. A. Reisenleiter, Army Research Institute, the examinee's score is based on the sum of the individual item discriminant weights as opposed to one for a correct response and zero for an incorrect one.

---

[1]The plug-in disciminant function was selected rather than a more, traditional approach, such as Fisher's (1936) linear discriminant function, because its statistical assumptions fit nicely within the framework of the problem addressed in this study, and the mathematical computations necessary to apply the plug-in technique are much shorter and simpler than those needed for more traditional approaches. While Fisher's technique, which is based on log-likelihood ratios, provides a useful tool for discriminating between populations, it may be quite unsuitable for allocating a particular subject to one of two populations when the underlying distribution is far from multivariate normal.

Purpose

The purpose of this effort was to determine whether or not an application of the plug-in discriminant function could significantly improve the overall predictive validity of the scores on a diagnostic criterion-referenced test.

METHOD

Test

The test employed in this research was the criterion-referenced, Basic Mechanical Procedures (BMP) test, which was developed at the Navy Personnel Research and Development Center to diagnose individual deficiencies within the boiler technician rating (BT) (Laabs, Harris, & Pickering, 1977). The BMP is keyed to the 14 instructional modules listed in Table 1, and has a prescribed administration time of 90 minutes. The response vectors and validity data for 200 BTs that were used in the construction and validation of the BMP test were available.

Table 1

Basic Skills and Knowledges Modules

Module	Title
1	Metal Fasteners, Hand Tools
2	Pipes, Tubings, Fittings
3	Packing, Gaskets, Insulation
4	Valves
5	Bearings, Lubrication
6	Pumps
7	Precision Measurement Instruments, Technical Manuals
8	Heat Properties, Heat Exchangers
9	Indicating Devices
10	Turbines, Couplings, Gears
11	Strainers, Purifiers
12	Low Pressure Air System and Compressor
13	Oil Pollution
14	Planned Maintenance System

Sample

The sample consisted of 200 BTs assigned to the shore-based Propulsion Engineering School, Service School Command, Great Lakes. Half of them--the pre-instruction group--were entering the modularized, self-paced curriculum, and half--the post-instruction group--had completed it. For each group, response data from 25 students randomly selected were put aside for cross-validation purposes.

Plug-in Discriminant Analysis

In this procedure, the items to be analyzed are assumed to be independent binary-random variables with possible values of 1 and 0. Moreover, it is necessary to maintain a minimum of 10 subjects for every item included in this particular application of the discriminant analysis to assure stable statistical estimates of the actual or population discriminant weights. Granting these assumptions are met by the BMP test modules, the model for the plug-in discriminant function is computed as follows:

$$D_{jI} = \sum_{i=1}^{I} [X_{ji} \text{ Log}_e \frac{\Psi_i^2}{\Psi_i^1} + (1 - X_{ji}) \text{ Log}_e \frac{1 - \Psi_i^2}{1 - \Psi_i^1}]$$

Where D_{jI} = The discriminant score for person j on I items.

Log_e = A logarithm with e as a base.

X_{ji} = Response of individual j to item i.

Ψ_i^1 = The probability of an examinee from group 1 answering item i correctly. Ψ_i is estimated, in this study, as the proportion of correct responses within the pre-instruction group (1) on item i.

Ψ_i^2 = The probability of an examinee from group 2 answering item i correctly. Ψ_i^2 is estimated, in this study, as the proportion of correct responses within the post-instruction group (2) on item i.

To illustrate the basic mechanical operation of this formula, consider the following numerical example for a single item module (I = 1), where there are five students in each instruction group. Suppose only one of the five pre-instruction members answers our hypothetical one-item test correctly (Ψ_1^1 = .20), compared to three of the five post-instruction members (Ψ_1^2 = .60). The resulting discriminant function would appear as follows:

$$D_{j1} = \mathrm{Log}_e \frac{.6}{.2} X_{j1} + \mathrm{Log}_e \frac{.4}{.8} (1 - X_{j1})$$

$$= 1.10X_{j1} - .69\,(1 - X_{j1}).$$

In this example, the plug-in discriminant assigns a positive item weight (i.e., discriminant score) to a correct outcome (X_{j1} = 1) on item 1 and a negative item weight to an incorrect outcome (X_{j1} = 0). Thus, an examinee who answered item one on our hypothetical test correctly would receive a discriminant score of 1.10, while one who answered item one incorrectly would receive a discriminant score of -.69.[2]

---

[2]In the event that Ψ_i^k equals unity or zero, the natural logarithm expression in the discriminant function yields an indeterminate solution (i.e., the logarithm to the base e is undefined at 0). For this reason, the following is recommended:

When $\Psi_i^k = 1$ set equal to $\dfrac{N_k - 1}{N_k}$

$\Psi_i^k = 0$ set equal to $\dfrac{1}{N_k}$

When N_k is the number of examinees in group k. This prevents the assignment of infinity as an item weight.

To demonstrate just how easy it is to compute item weights in this manner, a simple FORTRAN program, appearing in Appendix A, was created to generate item weights and subsequent discriminant scores. A sequence of information describing the necessary input parameters must be entered as part of the program before any data analysis can begin. This information includes (1) the total number of subjects and items to be processed, (2) the two $\hat{\Psi}$ parameter estimates for each item of the test, and (3) a binary-scored response vector for each subject. The output for each plug-in discriminant analysis includes (1) the output identification card to be used as a heading to your output, (2) the discriminant weights for each item, and (3) the discriminant scores for each subject. The program is presently limited to include a maximum of 200 subjects and 80 items. Of course, all of the variable size limitations can be altered by increasing array sizes within the program.

The FORTRAN program, using the 150-item response sets from the pre- and post-instruction BMP test answer sheets, was accessed to compute the discriminant coefficients for each of the items and subsequent discriminant scores for the examinees in the validation group. This analysis was repeated for each module in a similar manner.

Estimating Validity

To evaluate the validity of the criterion-referenced BMP test modules, a predictive evaluation approach was taken (see Laabs & Panell, 1978, or Panell & Laabs, 1979). This approach to the problem of estimating the validity of criterion-referenced tests followed a prescribed set of procedures that includes the establishment of a cutoff score for each module and the subsequent use of this criterion to classify students in the cross-validation sample. Specifically, if an examinee's conventional scores fell at or above the appropriate cutoff score, he was classified as a post-instruction group member. If his performance fell below the criterion, he was classified as a pre-instruction group member. This classification was compared to actual group membership in the validation sample and the

percentage of agreement was determined. It should be noted that this concept of validity necessarily assumes that the instructional material to which the test is being keyed is effective.

The estimates of the validity of the criterion-referenced BMP test modules using the conventional scoring method was compared to that using the discriminant scoring method. In the discriminant scoring method, the item weights within each module were computed using the developmental sample of 150 BTs, with the cutoff score in this instance equaling $\text{Log}_e\ \pi 1/\pi 2$, where $\pi 1$ and $\pi 2$ equal the proportion of pre- and post-instruction members in the cross-validation groups. Since the proportion of pre-instruction members equaled the proportion of post-instruction members in the validation sample, the cutoff criterion equaled $\text{Log}_e 1$ or zero for each module. Predictive validity was therefore estimated by using this new criterion to classify the examinees in the cross-validation sample as illustrated below.

Assign subject j to the pre-instruction group if D_{jI} is less than C and to the post-instruction group if D_{jI} is greater than or equal to C, where the cutoff point C equals zero.[3] This classification was compared to actual group membership in the cross-validation sample, and the percentage of agreement was determined.

---

[3]The cutoff criterion employed in the discriminant scoring procedure was chosen for use to minimize the total probability of misclassifying examinees in terms of group membership (Glick, 1973).

RESULTS AND DISCUSSION

Table 2 reports the estimated validity coefficients for 14 module subtests predicted from number-correct and discriminant scores (modules 4 and 6 did not meet the subject-to-item ratio requirement for discriminant scores). These results show that most of the subtests, under both scoring methods, demonstrated excellent classification ability. Each of the percent agreements was significant at conventional levels, as determined by a chi-square test with Yates correction for continuity (Appendices B and C contain the contingency tables and cross-validation data). Although the uniformly large classification percentages achieved under the number-correct scoring method indicated little room for improvement in individual subtest validity, an inspection of Table 2 reveals an increase in the overall validity of the discriminant scores as compared to the number-correct scores. An analysis was performed to confirm this observation by comparing the validity coefficients obtained under both scoring procedures across 12 modules--all but modules 4 and 6. The differences between validity coefficients for the 12 modules were then submitted to a Wilcoxen-pairs signed-rank test, which showed a significant improvement in the overall test validity ($Z = 5.39$, $p < .001$). Moreover, the differences between pairs of validity coefficients in the 5 modules having the poorest classification percentages (Nos. 2, 3, 5, 8, and 9) were analyzed, in a post hoc fashion, by a statistical test for correlated proportions; all but module 9 were significant at conventional levels.

The outcome of this study regarding the predictive validity of number-correct and discriminant scores on the BMP test modules was in the expected direction. Scores resulting from the discriminant weights improved the ability of this criterion-referenced

Table 2

Agreement in Classification of a Cross-Validation Sample
N = 50

Module	Title	Number-Correct Percent Agreement	Discriminant Weights Percent Agreement
1	Metal Fasteners, Hand Tools	.88	.88
2	Pipes, Tubings, Fittings	.78	.86
3	Packing, Gaskets, Insulation	.68	.82
4	Valves	.86	--[a]
5	Bearings, Lubrication	.74	.78
6	Pumps	.92	--[a]
7	Precision Measurement Instruments Technical Manuals	.86	.88
8	Heat Properties, Heat Exchangers	.72	.78
9	Indicating Devices	.78	.80
10	Turbines, Couplings, Gears	.80	.78
11	Strainers, Purifiers	.88	.88
12	Low Pressure Air System and Compressor	.88	.90
13	Oil Pollution	.86	.86
14	Planned Maintenance System	.90	.92

[a]An insufficient ratio of subjects to items existed in modules 4 and 6, thus violating one of the basic assumptions of the discrete discriminant model.

test to accurately classify students in the cross-validation sample as members of either the pre- or post-instruction group. It is particularly important to note that, although this test had a very high validity content before the application of discriminant weights, the differences between the classification percentages for the scoring procedures favored the discriminant application in all but one of the modules. A small negative difference was recorded for module 10. Thus, it would appear that an application of the "plug-in" discriminant function did improve the overall validity of this criterion-referenced test.

CONCLUSION

The results of this effort demonstrate that discriminant coefficients used as item weights for scoring purposes can significantly enhance the BMP test's capacity to correctly discriminate between those students who need additional instruction and those who do not. This scoring procedure resulted in an increase in predictive validity by maximizing through differential item weighting the differing validity levels of the individual items. The only limitation of this procedure is the 10 to 1 subject-to-item ratio that must exist before using this application of the discriminant model. In addition to the gains in psychometric qualities displayed by this application of the plug-in discriminant function, the following properties should also be included as assets of this procedure:

1. The discriminant function would appear to be sound by analogy to statistical theory and would utilize well known item statistics.

2. The discriminant function allows the items to be scaled along a discrimination continuum with meaningful end points.

3. The discriminant function provides an index of the usefulness of the items for discriminating between pre- and post-instruction member's test scores. Such a determination of item weights would be easy to program, as is demonstrated by the example appearing in Appendix A.

4. The discriminant function should prove especially helpful in improving the overall validity of tests whose items are not as valid to begin with as the BMP test.

RECOMMENDATION

In view of these properties, the plug-in discriminant functions should be considered as an alternative procedure to the more conventional number-correct procedure for determining item scores. This grading method for the BMP test, coupled with an adequate item development procedure such as the one outlined in Laabs and Panell (1978) and Panell and Laabs (1979), provides the researcher with an excellent means for maximizing item information on criterion-referenced tests.

REFERENCES

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Coombs, C. H., Milholland, J. E., & Womer, F. The assessment of partial knowledge. Educational and Psychological Measurement, 1956, 16, 13-37.

Elvers, E. A comparison of the two discriminants--"plug-in" and likelihood ratio. Stockholm: Research Institute of National Defense, August 1977.

Fisher, R. A. The use of multiple measurements in taxonomic problems. Annuals of Eugenics, 1936, 7, 179-188.

Glick, N. Sample-based multinominal classification. Biometrics, 1973, 29, 241-256.

Guttman, L. The quantification of a class of attributes: A theory and method of scale construction, In P. Horst, et al. (Eds.). The prediction of personal adjustment. New York: Social Science Research Council, 1941.

Hambleton, R. K, & Novick, M. R. Toward an integretation of theory and method for criterion-referenced test. Journal of Educational Measurement, 1973, 10, 159-170.

Hayes, W. L. Statistics for the social sciences. New York: Holt, Rhinehart, and Winston, 1973.

Laabs, G. J., & Panell, R. C. Construction of a criterion-referenced, diagnostic test for boiler technicians (NPRDC Tech. Note 78-14). San Diego: Navy Personnel Research and Development Center, May 1978.

Laabs, G. J., Harris, H. T., & Pickering, E. J. A personnel readiness training program: Operation and maintenance of the 1200 PSI Steam Propulsion Plant (NPRDC Tech. Rep. 77-36). San Diego: Navy Personnel Research and Development Center, June 1977. (AD-A042 033)

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison, Wesley, 1968.

Panell, R. C., & Laabs, G. J. Construction of a criterion-referenced diagnostic test for an individualized instruction program. Journal of Applied Psychology, 1979, 64, 255-261.

Patnaik, D., & Traub, R. E. Differential weighting by judged degree of correctness. Journal of Educational Measurement, 1973, 10, 281-286.

Popham, W. J., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Shuford, E. H., Albert, A., & Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 21-35.

Stanley, J., & Wong, M. D. Weighting, test-items and test-items options: An overview of the analytical and empirical literature. Educational and Psychological Measurement, 1970, 30, 21-35.

APPENDIX A

PROGRAM FOR PLUG-IN DISCRIMINANT FUNCTION ANALYSIS

```
C PROGRAM LANGUAGE = FORTRAN IV
C  PROGRAM WRITTEN FOR UNIVAC 1110
C
C MAXIMUM CAPABILITIES ARE-
C  UP TO 200 EXAMINEES
C  NUMBER OF SS READ AS NN ON DATA PARAMETER CARD.
C  UP TO 80 ITEMS
C  NUMBER OF ITEMS READ AS II ON DATA PARAMETER CARD.
C  NUMBER OF SS OR ITEMS MAY BE INCREASED IF NECESSARY BY CHANG-
C  ING DIMENSIONS.
C
C
C THIS ANALYSIS REQUIRES THAT THE PARAMETER ESTIMATES ARE 0<PSI<1
C AND THE SUBJECTS' VECTOR SCORES ARE GRADED IN A DICHOTOMOUS
C MANNER(0-1).
C
C SUBMIT CARDS IN FOLLOWING ORDER:
C
C  PROGRAM CARDS
C
C YOUR IDENTIFICATION CARD. USE ANY OF THE 80 COLS. FOR INFORMA-
C TION THAT YOU WISH TO APPEAR AS A HEADING FOR YOUR OUTPUT.
C
C THE NEXT CARD IS THE DATA PARAMETER CARD AND SHOULD BE PUNCHED
C IN THE FOLLOWING MANNER:
C COL. 1-3 THE NUMBER OF SUBJECTS(UP TO 200).
C COL. 4-6 THE NUMBER OF ITEMS(UP TO 50).
C
C THE NEXT CARDS ARE THE PSI PARAMETER ESTIMATE CARDS AND SHOULD
C BE PUNCHED IN THE FOLLOWING MANNER:
C COL.1-4 PARAMETER ESTIMATE FOR GROUP1 ITEM1.
C COL.5 BLANK
C COL.6-9 PARAMETER ESTIMATE FOR GROUP1 ITEM2.
C COL.10 BLANK
C CONTINUE IN THIS MANNER FOR EACH OF THE ESTIMATES IN PSI1,THER
C BEGINNING ON A NEW CARD REPEAT THIS PROCEDURE FOR PSI2.
C
C THE NEXT COLLECTION OF CARDS IN THE DATA CARDS, WHERE EACH SHOULD
C BE PUNCHED IN THE FOLLOWING MANNER:
C COL.1-80(IF NECESSARY) THE DICHOTOMOUSLY SCORED(0-1) RESPONSE
C VECTOR
C
      DIMENSION ISCOR(200,80),PSI(2,80),DISCOR(200),DISWT(80,2),
     1 IDENT(20)
C
      READ(5,11) (IDENT(M), M=1,20)
   11 FORMAT(20A4)
      WRITE(6,12) (IDENT(M), M=1,20)
   12 FORMAT(' ',20A4,2(/))
C
C  READING DATA PARAMETERS.
      READ(5,13) NN,II
   13 FORMAT(2I3)
C
C  READING PSI PARAMETER ESTIMATES.
      DO 8 IT=1,2
        READ(5,14) (PSI(IT,I), I=1,II)
   14   FORMAT(16(F4.3,1X))
    8 CONTINUE
C
```

```
C  READING SUBJECTS' VECTOR SCORES.
      DO 9 J=1,NN
        READ(5,15) (ISCOR(J,K), K=1,II)
 15     FORMAT(80I1)
  9   CONTINUE
C
C  CLEARING DISCRIMINANT ARRAYS TO ZERO.
      DO 10 IN=1,NN
        DISCOR(IN)=0.0
 10   CONTINUE
C
C  COMPUTE ITEM DISCRIMINANT COEFFICIENTS.
      DO 20 I=1,II
        DISWT(I,1)=ALOG(PSI(2,I)/PSI(1,I))
        DISWT(I,2)=ALOG((1.0-PSI(2,I))/(1.0-PSI(1,I)))
 20   CONTINUE
C
C  PRINT DISCRIMINANT COEFFICIENTS.
      WRITE(6,16)
 16   FORMAT(4X,'ITEM WEIGHTS',2(/))
C
      DO 30 IC=1,II
        WRITE(6,17) (DISWT(IC,JJ), JJ=1,2)
 17     FORMAT(1X,F7.3,2X,F7.3)
 30   CONTINUE
C
C  COMPUTE DISCRIMINANT SCORE.
      DO 40 N=1,NN
        DO 50 I=1,II
          D=DISWT(I,1)*ISCOR(N,I)+
     1      DISWT(I,2)*(1-ISCOR(N,I))
          DISCOR(N)=DISCOR(N)+D
 50     CONTINUE
 40   CONTINUE
C
C  PRINT DISCRIMINANT SCORES.
      WRITE(6,18)
 18   FORMAT(2(/),'  DISCRIMINANT SCORES',2(/))
C
      WRITE(6,19) (KK,DISCOR(KK), KK=1,NN)
 19   FORMAT(1X,I1,5X,F8.3)
     STOP
     END
```

APPENDIX B

NUMBER-CORRECT: CONTINGENCY TABLES AND CROSS-VALIDATION DATA

NUMBER-CORRECT: CONTINGENCY TABLES AND CROSS-VALIDATION DATA

Module 1 Criterion = 7/8 % Agreement = 88 $\underline{X}^2 = 26.09$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group Membership	Pre	21	4
	Post	2	23

M dule 2 Criterion = 3/6 % Agreement = 78 $\underline{X}^2 = 16.77$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group Membership	Pre	14	11
	Post	0	25

Module 3 Criterion = 3/5 % Agreement = 68 $\underline{X}^2 = 5.12$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Acutal Group Membership	Pre	17	8
	Post	8	17

Module 4 Criterion = 11/17 % Agreement = 86 $\underline{X}^2 = 23.16$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	21	4
	Post	3	22

Module 5 Criterion = 4/8 % Agreement = 74 $\underline{X}^2 = 9.96$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	19	6
	Post	7	18

Module 6 Criterion = 9/12 % Agreement = 92 $\underline{X}^2 = 32.31$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	24	1
	Post	3	22

Module 7 Criterion = 7/10 % Agreement = 86 $\underline{X}^2 = 23.16$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	21	4
	Post	3	22

Module 8 Criterion = 2/4 % Agreement = 72 $\underline{X}^2 = 8.21$, $\underline{p} < .01$

		Diagonosed Group Membership	
		Pre	Post
Actual Group	Pre	16	9
Membership	Post	5	20

Module 9 Criterion = 5/7 % Agreement = 78 $\underline{X}^2 = 13.54$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	19	6
Membership	Post	5	20

Module 10 Criterion = 5/7 % Agreement = 80 $\underline{X}^2 = 16.64$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	17	8
Membership	Post	2	23

Module 11 Criterion = 4/6 % Agreement = 88 $\underline{X}^2 = 26.60$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	20	5
Membership	Post	1	24

Module 12 Criterion = 6/8 % Agreement = 88 $\underline{X}^2 = 26.09$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	21	4
	Post	2	23

Module 13 Criterion = 3/4 % Agreement = 86 $\underline{X}^2 = 24.08$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	19	6
	Post	1	24

Module 14 Criterion = 4/6 % Agreement = 90 $\underline{X}^2 = 29.30$, $\underline{p} < .01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	21	4
	Post	1	24

APPENDIX C

DISCRIMINANT: CONTINGENCY TABLES AND CROSS-VALIDATION DATA

DISCRIMINANT: CONTINGENCY TABLES AND CROSS-VALIDATION DATA

Module 1 Criterion = 0 % Agreement = 88 $\underline{X}^2 = 26.09$, $\underline{p}<.01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	21	4
Membership	Post	2	23

Module 2 Criterion = 0 % Agreement = 86 $\underline{X}^2 = 23.46$, $\underline{p}<.01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Post	20	5
Membership	Post	2	23

Module 3 Criterion = 0 % Agreement = 82 $\underline{X}^2 = 18.75$, $\underline{p}<.01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	23	2
Membership	Post	7	18

Module 5 Criterion = 0 % Agreement = 78 $\underline{X}^2 = 15.91$, $\underline{p}<.01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	21	4
Membership	Post	7	18

Module 7 Criterion = 0   % Agreement = 88 $\underline{X}^2 = 26.09$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	21	4
Membership	Post	2	23

Module 8 Criterion = 0   % Agreement = 78 $\underline{X}^2 = 13.54$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	19	6
Membership	Post	5	20

Module 9 Criterion = 0   % Agreement = 80 $\underline{X}^2 = 15.78$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	19	6
Membership	Post	4	21

Module 10 Criterion = 0   % Agreement = 78 $\underline{X}^2 = 14.08$, $\underline{p} < .01$

		Diagnosed Group Membership	
		Pre	Post
Actual Group	Pre	17	8
Membership	Post	3	22

Module 11 Criterion = 0 % Agreement = 88 $X^2 = 26.6$, $p<.01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	20	5
	Post	1	24

Module 12 Criterion = 0 % Agreement = 90 $X^2 = 29.30$, $p<.01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	21	4
	Post	1	24

Module 13 Criterion = 0 % Agreement = 86 $X^2 = 24.08$, $p<.01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	19	6
	Post	1	24

Module 14 Criterion = 0 % Agreement = 92 $X^2 = 32.00$, $p<.01$

		Diagnosed Group Membership: Pre	Post
Actual Group Membership	Pre	23	2
	Post	2	23

DATE
FILMED

7-8